LQCD Optimized Clusters
(and cluster work at JLab)

Chip Watson
Jefferson Lab

Motivation

- Moore's Law delivers increases in processor price performance of the order of 60% per year
- A high volume market has driven the cost of CPUs and components extremely low, with newer components available every few months, allowing increased capability each year at constant investment
- Home video gaming has encouraged the development of multi-media extensions; these small vector processors on commodity processors deliver super-scalar performance, exceeding 4 Gflops sustained (single precision, on a very small problem) on a 1.7 GHz Pentium 4 – scaling this to a cluster is the challenge!
- Cluster interconnects are maturing, allowing ever larger clusters to be constructed from semi-commodity parts

SciDAC Prototype Clusters

The SciDAC project is funding a sequence of cluster prototypes which allow us to track industry developments and trends, which also deploying critical compute resources.

Myrinet + Pentium 4

- 48 dual 2.0 GHz P4 at FNAL (Spring 2002)
- 128 single 2.0 GHz P4 at JLab (Summer 2002)
- 128 dual 2.4 GHz P4 at FNAL (Fall 2002)

Alternative cluster designs are now emerging…

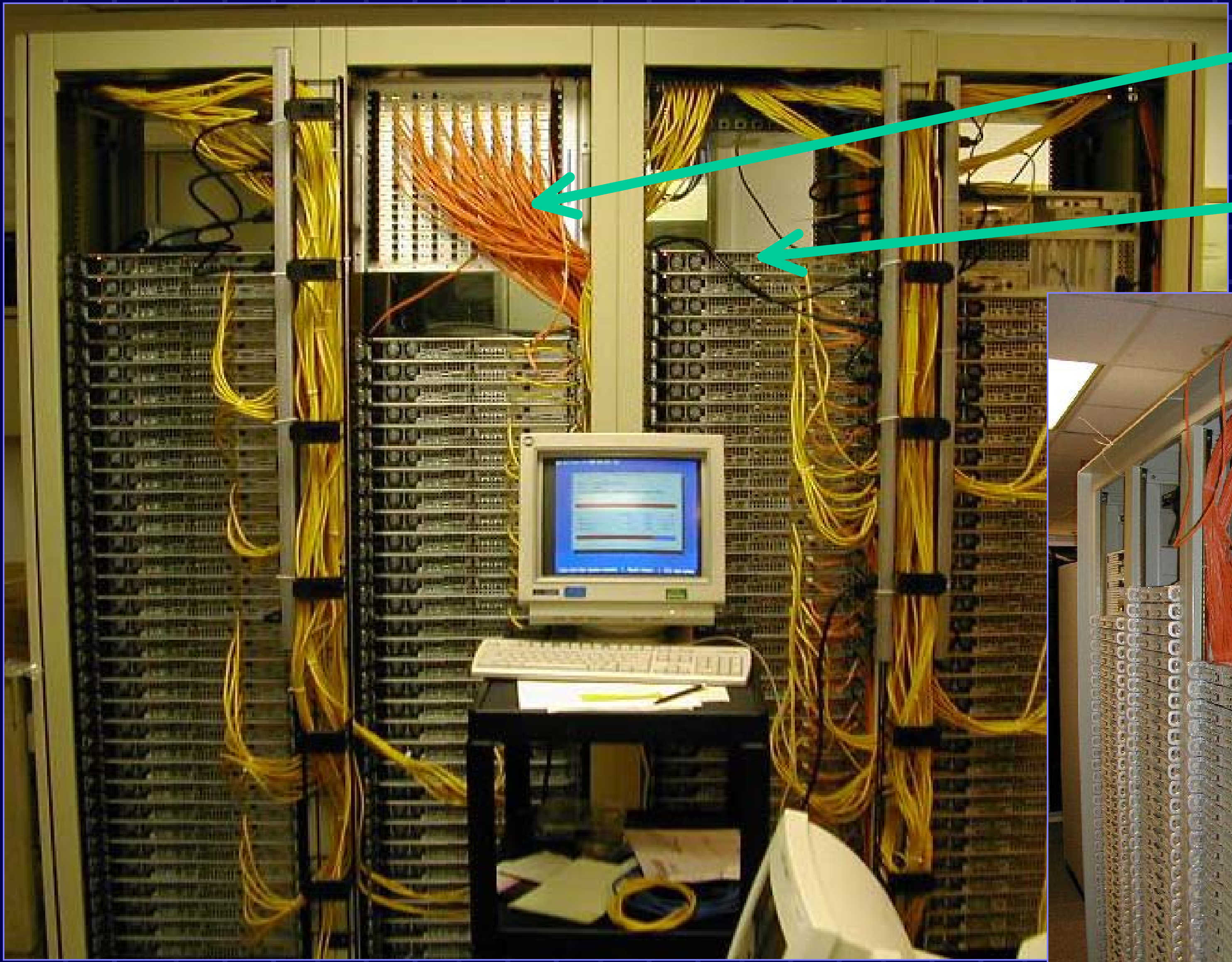
Gigabit Ethernet Mesh + Pentium 4

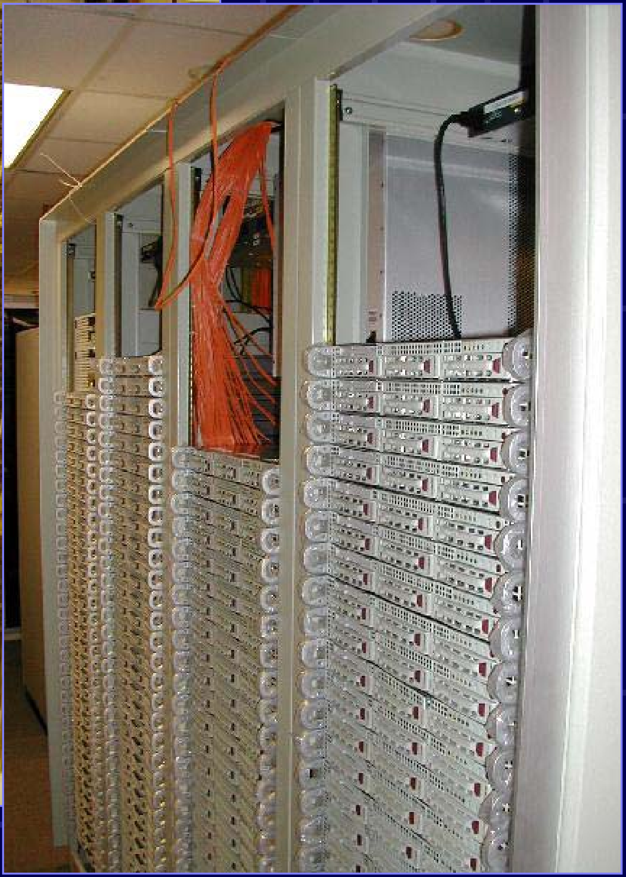
- 256 (8x8x4) single 2.66-2.8 GHz P4 at JLab (planned, Spring 2003)

Additional Technology Evaluations at FNAL for Summer 2003

- Itanium 2
- AMD Opteron
- IBM PowerPC 970

128 Node Cluster @ JLab

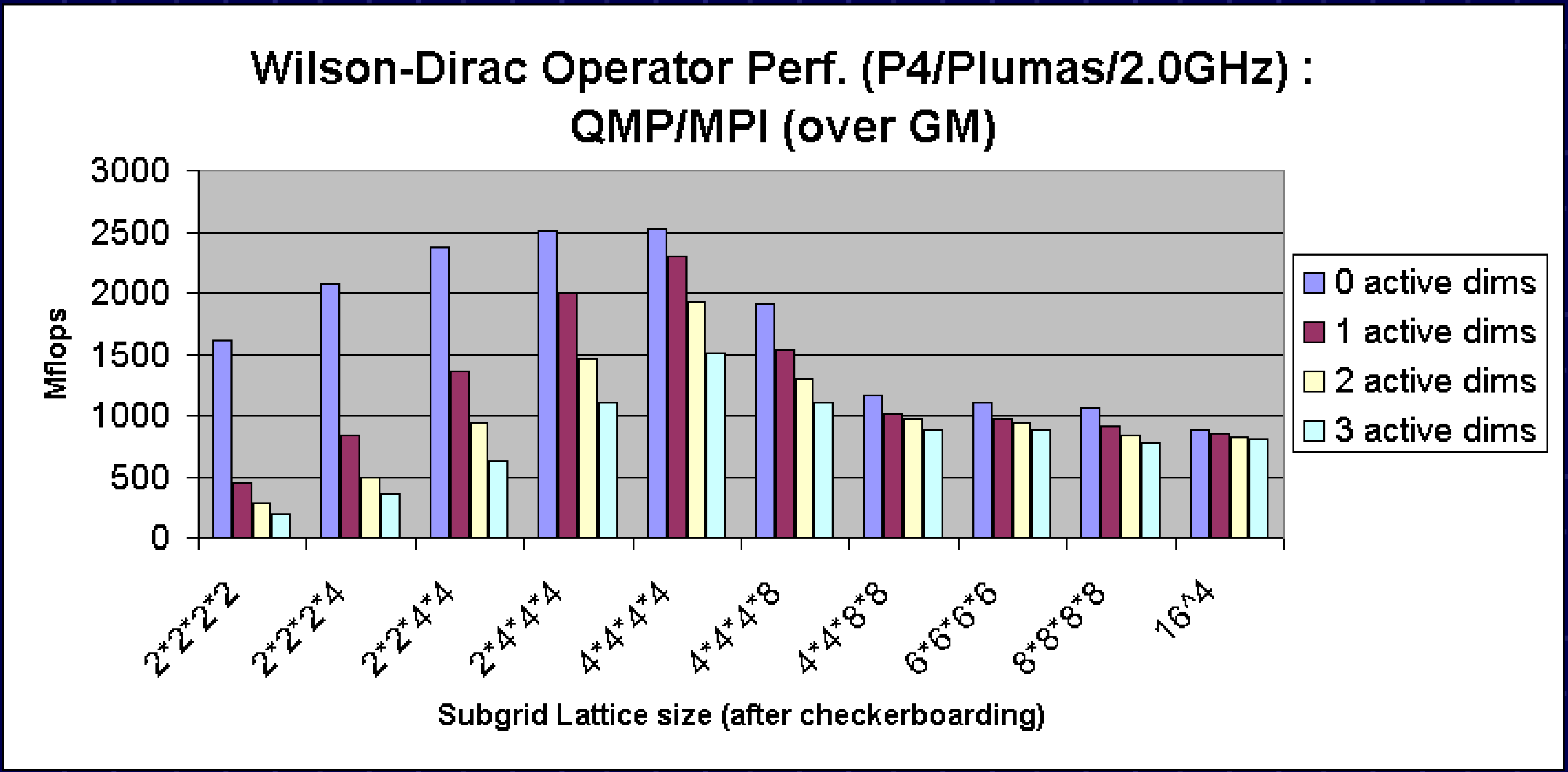




Myrinet

2 GHz P4
1U, 256Mb

2002 Cluster Performance



Cluster Strategy

Commodity Clusters allow us to take advantage of the latest developments in processor design, memory sub-systems, and interconnect technology

- CPU's accelerate at ~ 60% / year (Moore's Law)
- Memory speed generally advances less rapidly and with fewer discrete steps, ~ 40% / year

=>Performance ratio of in-cache to out-of-cache is growing

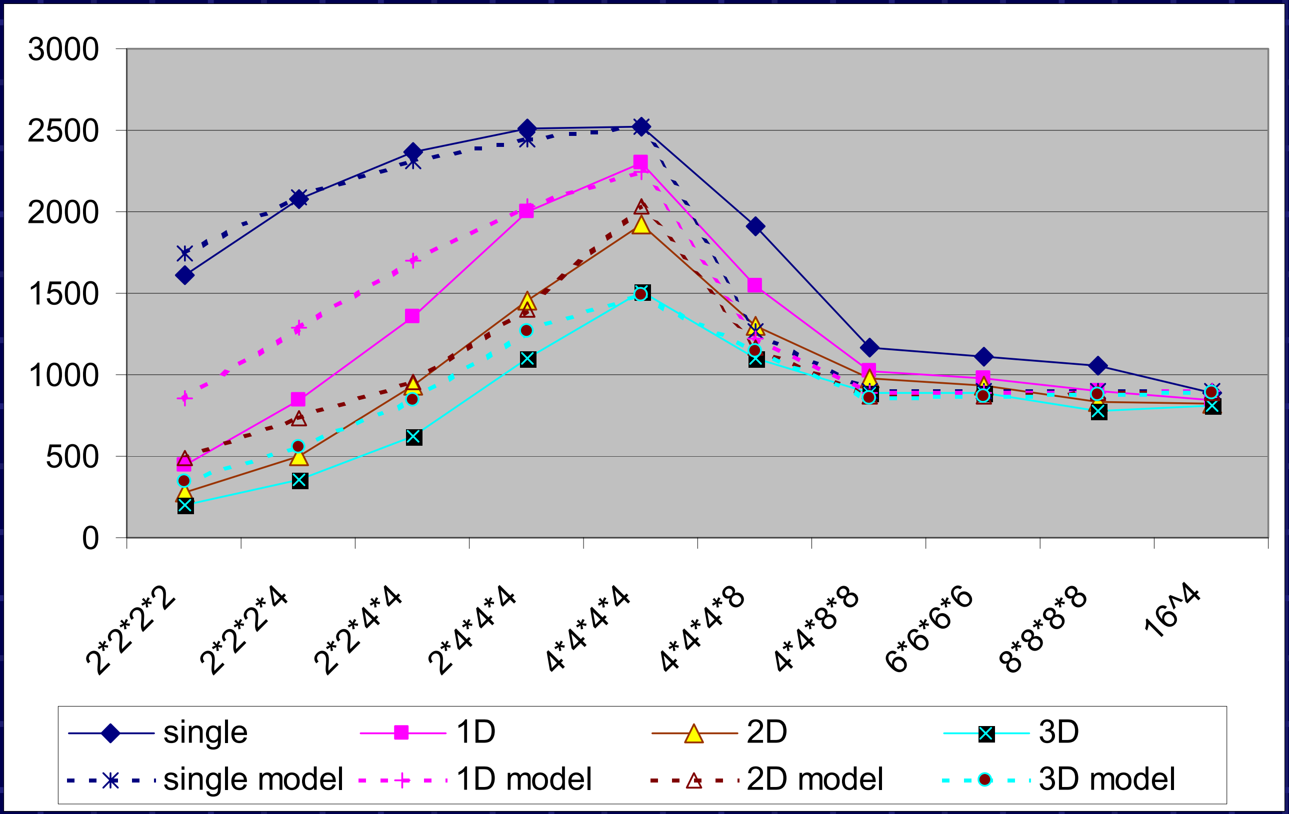
- Implications:

  Want to run as many applications in cache as possible (2x - 4x gain)

  => a large cluster used for single application

  => very high message rates (> 10 kHz !)
- Interconnects track external bus speeds, and server class motherboards will support processor evolutions for the next 2-3 years (multiple PCI-X busses)

Modeling Cluster Performance



- Model includes CPU in- and out-of-cache performance, PCI and link bandwidth, latency, etc.
- Moderately simple model predicts cluster performance pretty well.

Cluster Scalability Today

- PCI 32/33 runs out of steam at around 128 nodes today, and at 0.25 Tflops, PCI 64/66 or PCI-X becomes more cost effective (higher cost, higher efficiency)
- Single box systems with high bandwidth capability (533 MHz memory, PCI-X) cost ~$1600 and deliver ~ 1.3 Gflops, or $1.25 / Mflops (out of cache).
- High performance network costs are significant, $1300 / node for myrinet, yielding, for a 128 node cluster, $2.4 / Mflops (includes network overhead; less if problem is cache resident)
- Myrinet 2000 is capable of ~ 400 MB/s (200 each way)
  - This bandwidth would support clusters of up to 512 cpu's with good efficiency on lattice sizes of high interest today (24^3 x 32)
  - A cluster of this size could run the problem in cache, with each node delivering ~2.2 Gflops, or $1.6 / Mflops at 0.6 Teraflops (estimate based upon extrapolations from preliminary measurements on our cluster)

What about GigE?

- GigE switches (at hundreds of ports) make the network cost of large switched gigE clusters almost as high as myrinet, with lower bandwidth and higher latency (dead end?)
- GigE Mesh: 8 GigE links @ 800 Mbits / link each way

  (on 2 PCI-X busses @ 50% utilization, aggregate is not measured yet)

  delivers enough bandwidth for 8 Gflops sustained / box
  (e.g. four 2 Gflops processors; model result)
- Need efficient user space code:
  - Each usec of message start overhead corresponds to 5%-10% in performance
  - for \$2M machine, worth 6 person-months to optimize for 1 usec!

2003 JLab gigE Mesh Machine

Preliminary studies indicate that gigE is viable today:

- Network cost for 3D mesh = $500/node

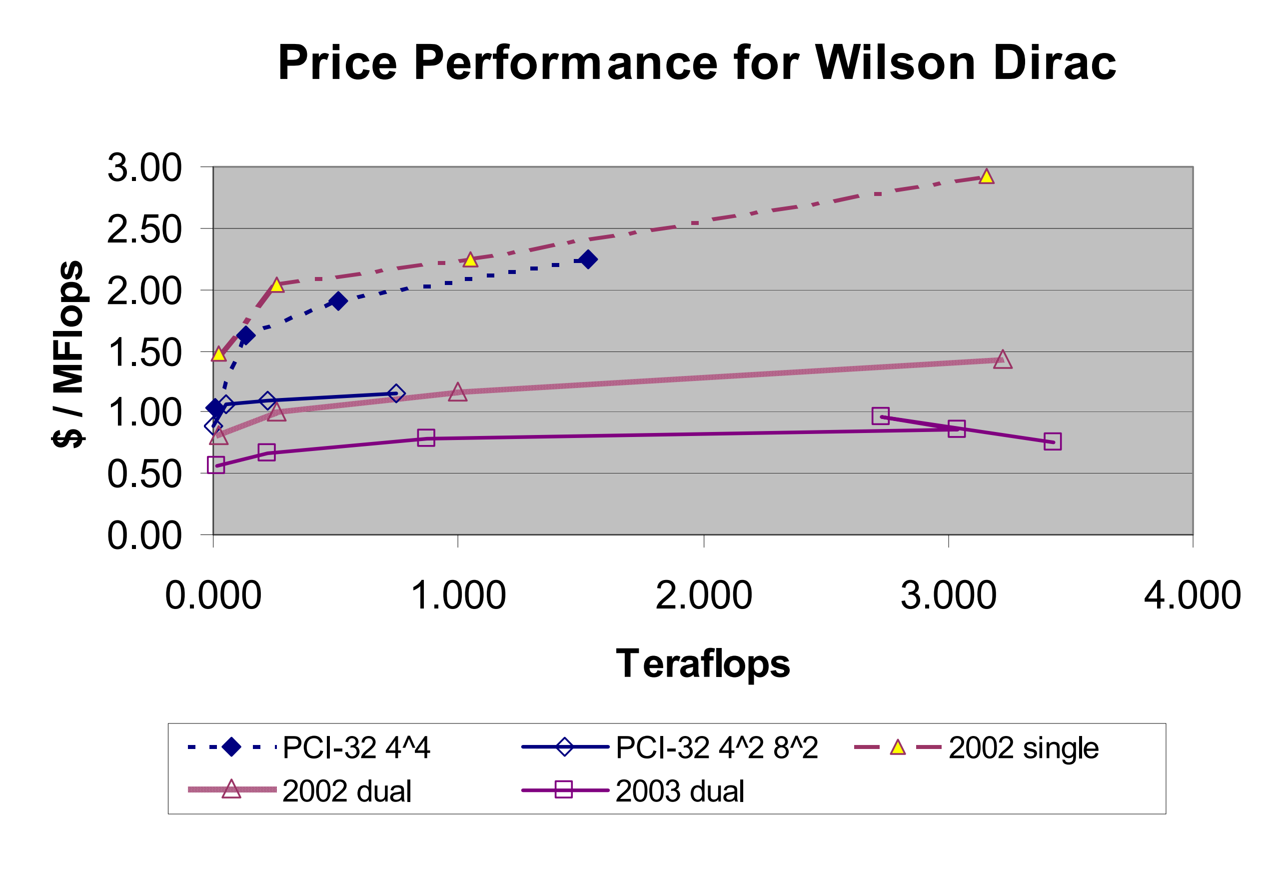
  (vs $1300 myrinet at 128 nodes, $1500 at 256 nodes, $1700 at 512)
- Bandwidth across 3 simultaneous transfers should exceed myrinet bandwidth (~ 6 transfers possible for domain wall)
- Latency for user space gigE driver should be comparable to or better than myrinet GM driver (below 10 microseconds)

Note: Andrew Pochinsky at MIT is working with JLab to develop a fast driver & QMP implementation for the Intel dual gigE card / chip.

Performance Extrapolations

- Cluster performance depends upon many factors:
  - Lattice size (bigger is more efficient for network, smaller allows faster processing in cache)
  - Processor speed
  - Memory bandwidth (effects efficiency of N^{th} processor)
  - Cache size
  - Link bandwidth
  - Link latency (mostly for global sums)
- Assumptions
  - Moore’s Law (60% processor improvement per year)
  - 2X step changes in link about every 2 years, achieving 50% of bus bandwidth (PCI-X, PCI-2X, Infiniband, …)
  - Quad processor servers become “commodity” by FY04

Future Clusters

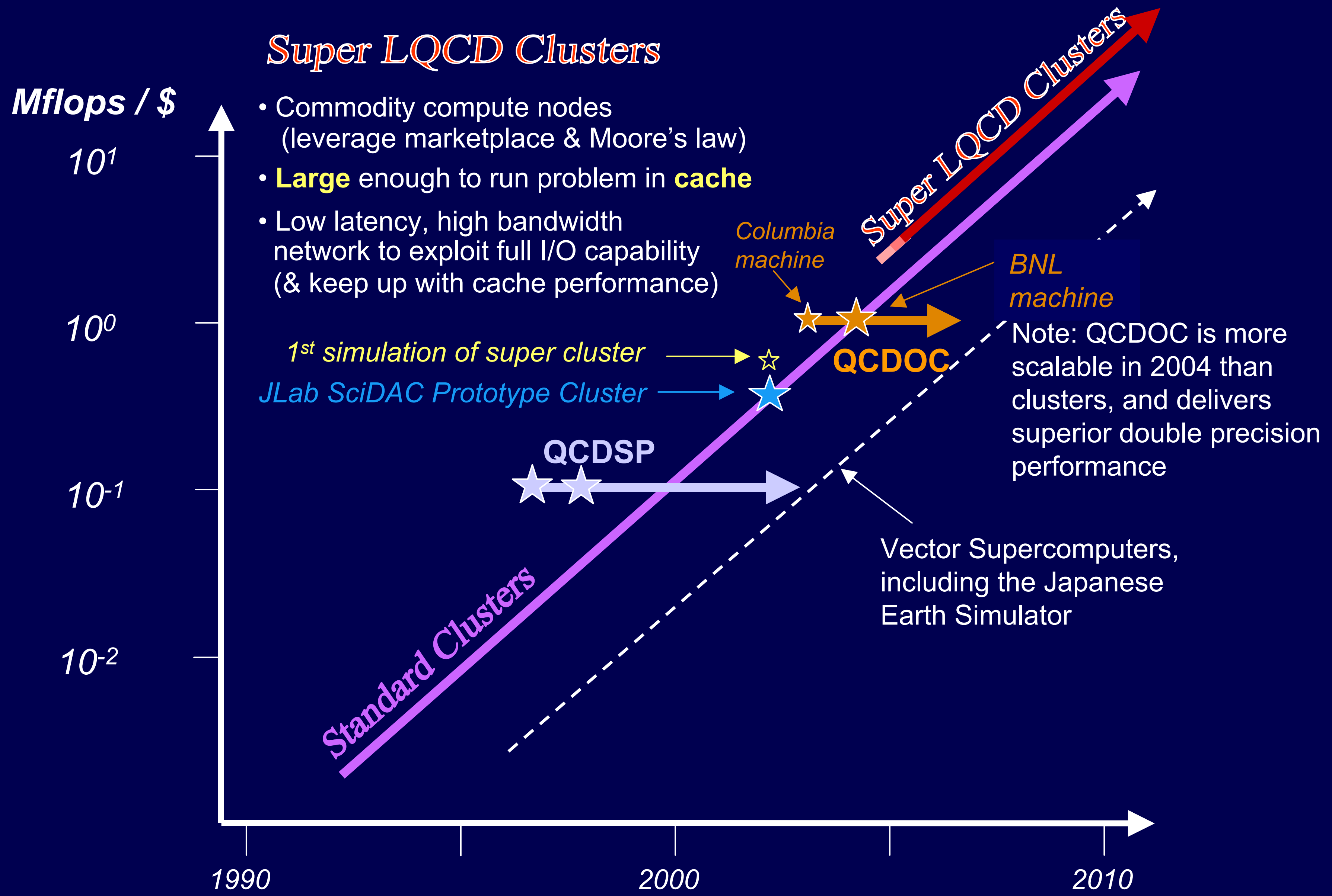


Assumption: lattice kernels running in cache.

(Figures are for late calendar year machines.)

For single precision, clusters could fall below $1/Mflop within a year. Implication: the 2 SciDAC architectures are complimentary in the near as well as mid term.

Performance per Dollar for Typical LQCD Applications



Four Year Plan

2003
- 256 node 8x8x4 gigabit ethernet mesh @ JLab
- (256?) node @ FNAL (alternate processor? – tbd)

2004
- Additional 256+ node prototypes (~ 0.5 Tflops sustained per cluster running in memory, not cache) at JLab and FNAL to explore latest options, possibly including custom NIC

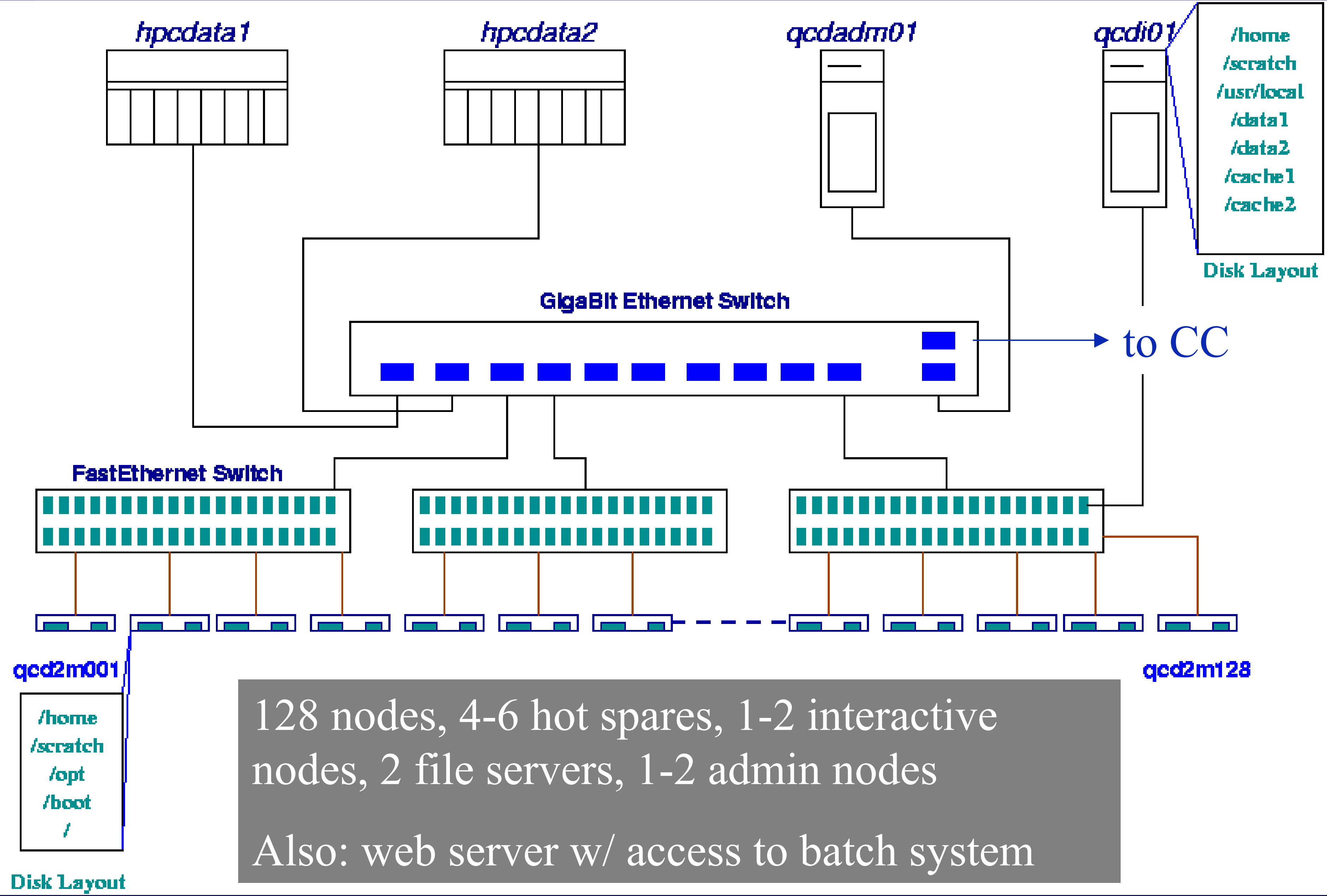
2005
- Large clusters of scale 3-4 Tflops
- Reference machine: 8x8x16 gigabit ethernet mesh, 4-way SMP Xeon (800 MHz FSB, 1.25 MB cache, 4 GHz dual processor core)

2006
- Large clusters of scale 5-6 TFlops

Cluster Usage at JLab

1. Get an account at JLab (fill out & sign a form; must be sponsored by one of the JLab staff in theory group for now) http://cc.jlab.org/docs/services/cue/accounts.html
2. Get account enabled on SciDAC cluster
3. ssh to interactive node
4. Move files from offsite to JLab
5. Edit batch script & qsub the script

Network, File System View



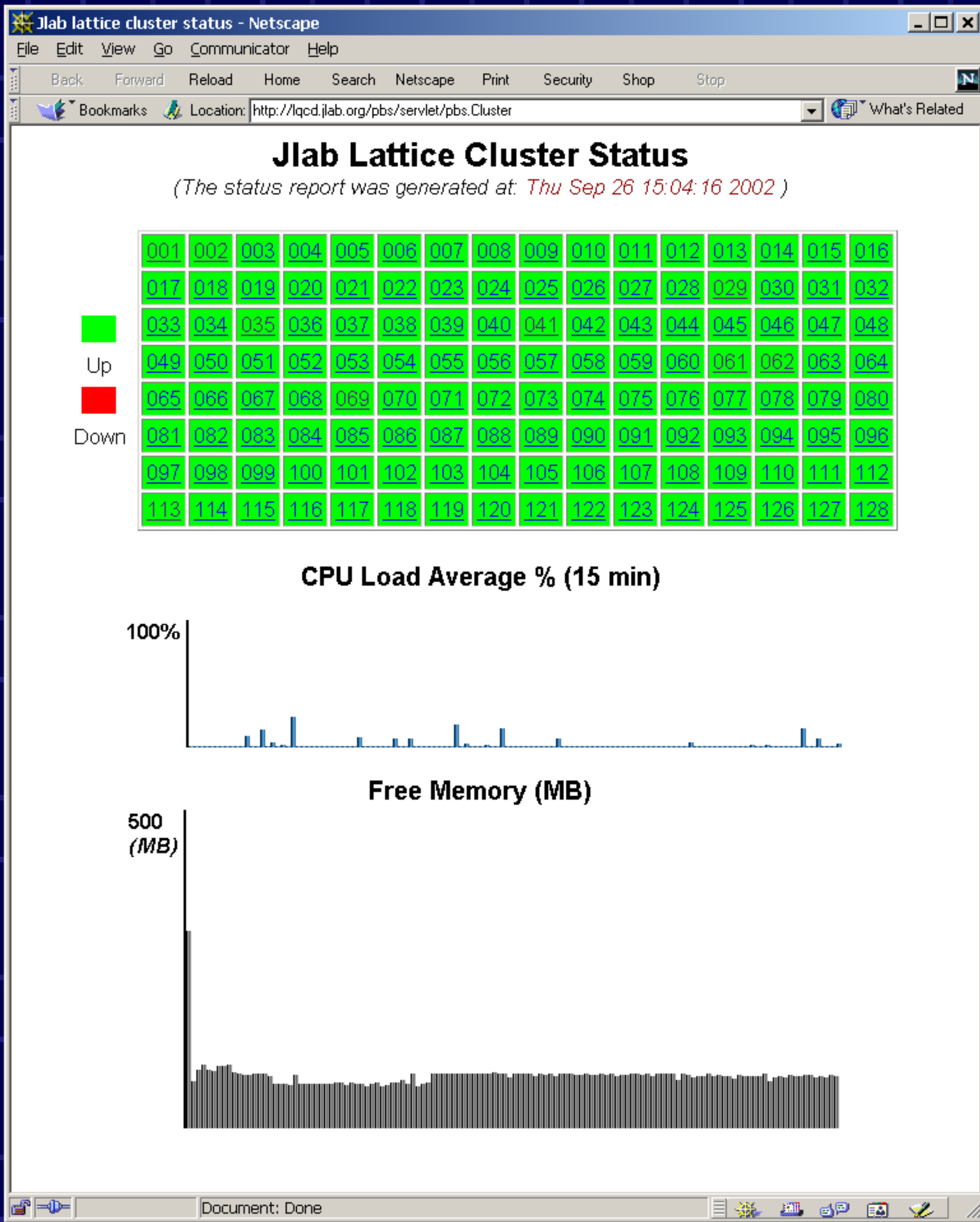
128 nodes, 4-6 hot spares, 1-2 interactive nodes, 2 file servers, 1-2 admin nodes

Also: web server w/ access to batch system

User Environment Info

- User's home directories are mounted on compute nodes (for now)
- 4+ Terabytes of disk are NOT mounted on compute nodes (so batch script must use rcp to get/put)
- Top level directories on big file servers must be created by sys-admin (e.g. on per project basis)
- Quotas are turned on (and large) on big file servers
- On 3 of the 4 terabytes, auto file migration to tape is done to maintain 5% free
- On "managed" file space, can pin, unpin, migrate to/from tape
- /home & large file servers are grid accessible: can use parallel file transfers to improve bandwidth to/from JLab

Long range goal: same user environment at all SciDAC sites. Work to define this environment has begun.

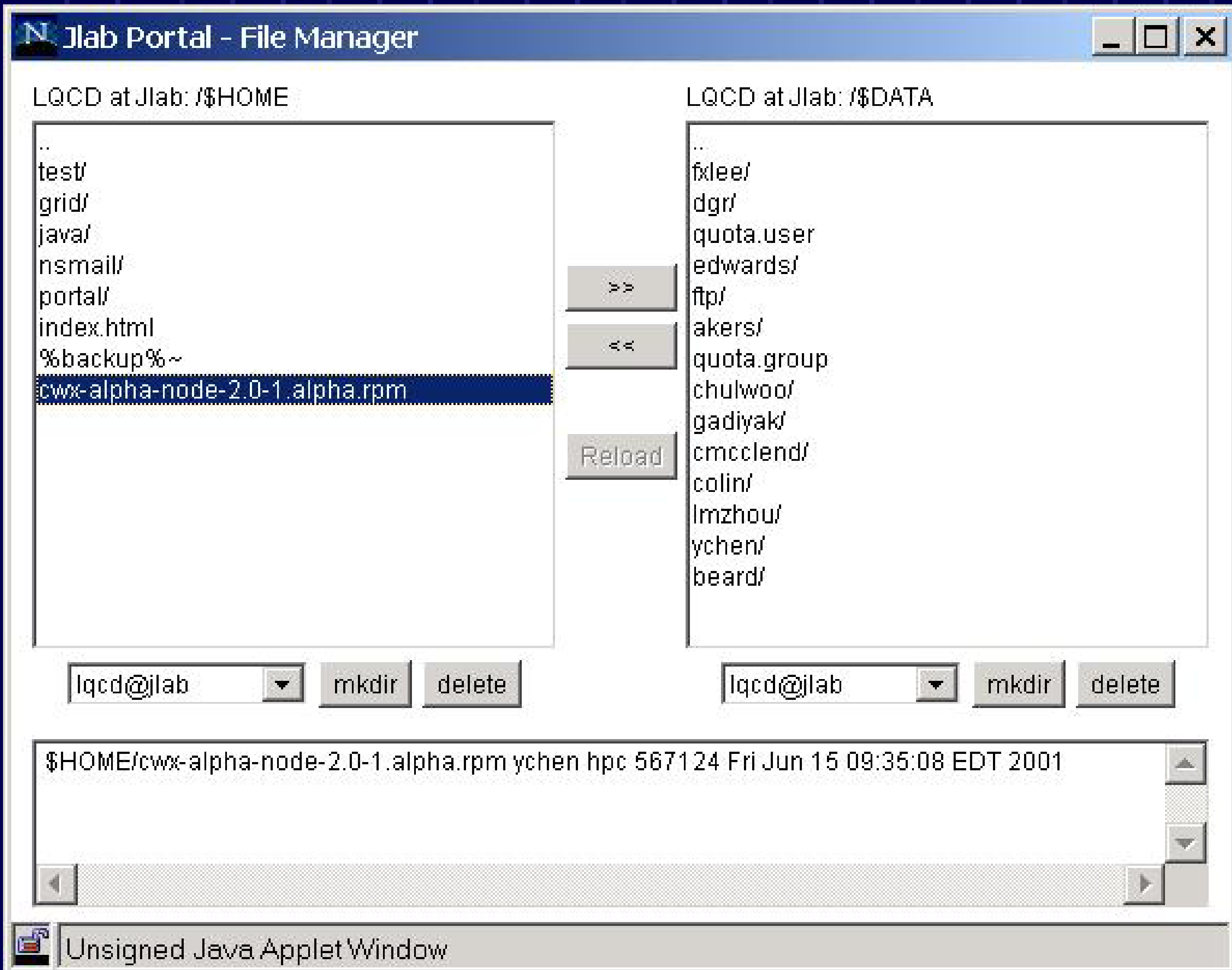


- Batch system status is web accessible
- Many different views available
- Long range goal: web services based computational grid, so users can submit to distributed LQCD facility

Secure Operations

- Remote secure operations require an X509 certificate. DOE runs a certificate authority to issue these certificates (DOE Science Grid), and as members of this SciDAC collaboration you are entitled to a certificate.
- Go to http://www.doegrids.org/ and select a virtual organization for requesting a certificate. LQCD is not yet a VO, so if you have no better option, you may choose PPDG (Particle Physics Data Grid), then "Request a user certificate"…
- Specify "Chip Watson" as the sponsor, and in the comments field put LQCD, and the name of your institution, and the name of someone I can ask who you are (to validate your request).

Data Grid File Manager

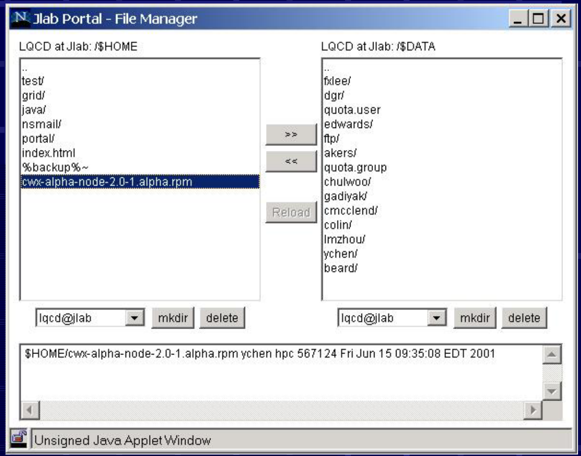


Starts via a web link

Requires an X509 certificate

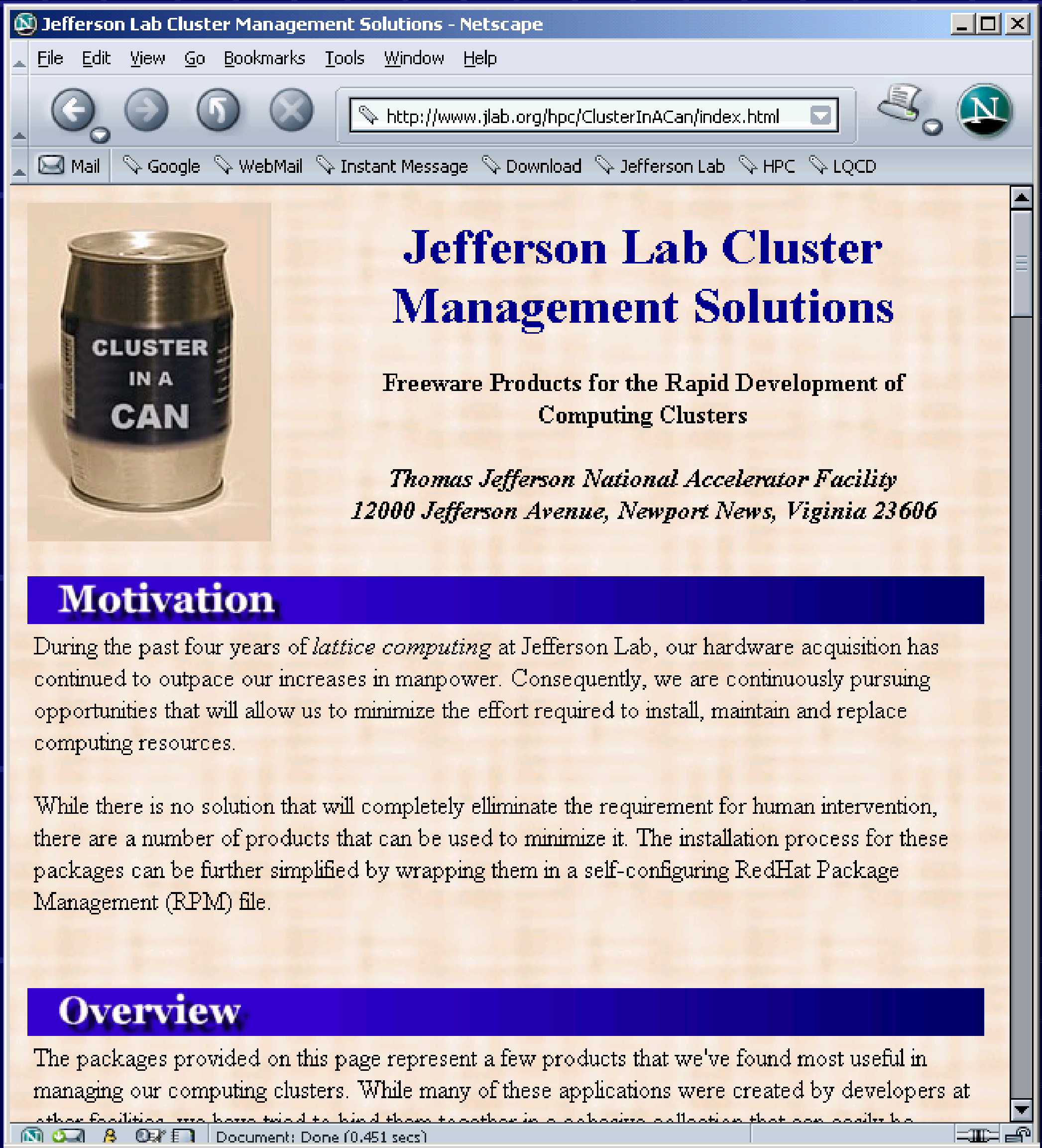
Growing functionality…

Capabilities (prototype)



- Browse contents of file system or replica catalog
  - Managed disk cache on data grid node
  - Unmanaged Local or Remote file system
  - Tertiary storage (eventually HRM)
- Move files between managed and unmanaged storage
  - Within a single data grid node
  - Between local file system and data grid node
  - Between data grid nodes (3rd party transfer)
- Status checks on long lived, asynchronous ops
- Migrate files from tape to disk

Support for Clusters



- We are currently working to package up all the tools we use to install & run our clusters, so that university groups can easily replicate a production environment.
- Assumes RedHat Linux, builds on open source software, tools
  - Install & config O/S, a few cluster monitoring & mgmt tools
  - Install batch (PBS)
  - Install web views
  - Install SRM (Storage Resource Manager)
- Will investigate distribution tools

For More Information

- New Lattice QCD Web Server / Home Page:

  http://www.lqcd.org/

- The Lattice Portal (currently presents JLab + MIT)

  http://lqcd.jlab.org/

- High Performance Computing at JLab

  http://www.jlab.org/hpc/